Condition Specific or Generic Preference-Based Measures in Oncology: the EORTC-8D or the EQ-5D

Paula Lorgelly[1], Brett Doble[1], Donna Rowen[2], John Brazier[2]

[1] Centre for Health Economics, Faculty of Business and Economics, Monash University
[2] Health Economics and Decision Science, School of Health and Related Research, University of Sheffield

Abstract

Background: It has been argued that generic HRQoL measures are not sensitive to certain disease specific improvements; condition-specific preference-based measures (CSPBMs) may offer a better alternative.

Aim: To assess the validity, responsiveness and sensitivity of a cancer specific preference based measure, the EORTC-8D, relative to the EQ-5D-3L.

Data: Cancer 2015 is a large-scale prospective longitudinal population-based molecular study. It enrols cancer patients who are treatment naïve. All cancer tumour sites, except leukaemia, and all grades of cancer from localised through to metastasised are included. Patients are followed for three, six and twelve months. Patients completed the EORTC-QLQ-C30 and the EQ-5D-3L at baseline and each follow-up.

Method: The EORTC-8D (which is derived from the EORTC-QLQ-C30) is compared to a generic measure, the EQ-5D-3L, in a cohort of cancer patients. Initially a qualitative assessment of each instrument is undertaken, comparing domains given their maximal theoretical contributions to the final utility score. Construct validity is assessed using spearman rank correlations coefficients for both the domains and the utility scores. Agreement between the two measures at baseline is evaluated using a normalised Bland-Altman plot and the intra-class correlation coefficient. The sensitivity of each instrument to different demographic and disease indicators is assessed using t-tests and ANOVAs. QALYs are estimated using the area under the curve method, and the effect of demographics, disease and disease progression on QALYs are evaluated using a seemingly unrelated estimation approach and Chow test to compare regression coefficients.

Results: Complete case analysis of 1062 patients found that at baseline the EQ-5D-3L values are significantly lower than the EORTC-8D values (0.752 v 0.833, $p<0.001$). The correlation between the instruments was strong, and agreement between the instruments was good. The baseline utility values were found to be sensitive to a number of patient and disease characteristics, which did not differ across instruments. Mean generic QALYs (estimated using the EQ-5D-3L) were lower than condition-specific QALYs (estimated using the EORTC-8D), and this difference was found to be significant. A Chow test failed to identify specific drivers of this difference in terms of patient and disease characteristics.

Conclusion: When comparing a generic and condition-specific preference-based instrument, divergences are apparent in both baseline utility and in the estimated QALYs over time for cancer patients. Condition-specific baseline utility scores are higher, and there is also greater QALY gain when using the condition-specific measure. Such divergences in QALYs between the generic measure and the condition specific measure will be problematic in an economic evaluation if it means that the funding decision changes depending on the instrument.

1. Introduction

Cost utility analyses require preference-based measures (PBMs) of outcome. Until recently most PBMs of outcome (so called multi-attribute utility instruments, MAUIs) have been generic. The mostly commonly employed generic PBM is the EQ-5D, a measure which the National Institute for Care and Excellence (NICE) actively encourage.[1] While the use of the same measure across a range of diseases and conditions increases comparability (what NICE refer to as a need for consistency) when informing decisions, there have been criticisms (both from health economists and clinicians) that these generic measures are not sensitive to certain disease specific improvements.[2] Their use in some diseases may mean that important clinical and patient quality of life changes may be underestimated.

There are a number of ways in which health economists can introduce condition-specific elements into a cost utility analysis.[3] Often we undertake mapping, where we estimate the relationship between a condition-specific non-preference-based measure and a generic PBM and apply this to the data.[4] A more resource intensive alternative is to elicit preferences from patients for condition-specific vignettes describing a health state. That is to use preference elicitation methods like time trade-off, standard gamble or a discrete choice experiment within the study population. A third alternative which is growing in popularity is to develop condition specific preference based measures (CSPBMs). Such CSPBMs can be developed from first principles (determine what is important to a patient/sufferer, undertake valuation study, and design the instrument),[5] or one could modify (in many instances this means reduce) an existing non-preference=based measure (as was the case with the SF-6D [6]). These non-preference-based measures have already been developed for the condition, so arguably have already been validated and had their sensitivity assessed. An additional benefit of using existing non-preference-based measures is that clinicians can get information on quality of life and disease domains of interest to them, while health economists are able to estimate utilities.

Despite the apparent benefits of CSPBMs their use is limited, in part due to NICE's requirement for generic (EQ-5D) values. If they are to be more widely adopted then evidence of their performance will be required. Using oncology as a case study this paper seeks to assess the validity, responsiveness and sensitivity of a cancer specific preference-based measure, the EORTC-8D, relative to the EQ-5D. Like most instrument comparison papers, much of the analysis will focus on baseline utility estimates, however, given cost utility analyses focus on the change over time, cancer-specific and generic quality adjusted life years (QALYs) will also be compared.

The paper proceeds by first describing the dataset that will be utilised, and the range of methods employed to test validity, responsiveness and sensitivity of the instruments and resulting QALYs. Results are then presented and discussed, including potential policy implications.

2. Methods

2.1. Data

Cancer 2015 is a large-scale (Framingham-type) prospective longitudinal population-based molecular cohort study. It enrols treat naïve cancer patients, irrespective of the tumour site (except leukaemia) and at all stages of disease. Recruitment is staged and phase 1 targeted the enrolment of 1,000 patients from 5 hospitals in Victoria, Australia. It aims to test and implement a new model of cancer diagnosis and treatment, with a specific focus on integrating molecular pathology into routine cancer diagnosis. Essentially tumour samples are genotyped, and actionable mutations are identified in order to inform cancer treatment at an individual level (e.g. personalised medicine).

Upon consenting, tumour samples (biopsies) and blood are collected and a baseline questionnaire is completed. The questionnaire collects information on patient demographics, tumour site and stage, treatment intentions and includes three patient reported outcome measures (PROMS), the EORTC-QLQ-C30, EORTC-8D (as an instrument) and the EQ-5D-3L. The protocol is that patients are followed at six months and 12 months, unless deemed to have severe disease, at which point they are followed up at three months (in the expectation that they will not survive to six months). At follow-up they again complete the PROMs and information on treatment (including pharmaceuticals, chemotherapy, radiotherapy, and surgery) is entered into the database by a research nurse from patient records.

The genotyping and mutation testing is undertaken by using the TruSeq Amplicon Cancer Panel which tests for mutations in 48 cancer genes. This information is then added to the database. At enrolment patients are also asked to consent to have their Federal and State administrative health data linked. The system in Australia is such that pharmaceutical and medical services health care resource use is collected by the Commonwealth Government, and they release this on consent to researchers. At a State level hospitalisation data is collected and collated (for all hospitals, although there is more information on admissions to public hospitals as the primary purpose is funding), and the Victorian Department of Health is able to link information on emergency and daycase presentations and inpatient admissions to external datasets. This dataset is arguably the most comprehensive in existence, as it includes genomic, clinical, quality of life and resource use records for a range of cancer patients. The focus of this paper is the measurement of quality of life using PROMs.

The first patient was recruited to Cancer 2015 in November 2011, and as of September 2013 there were 1,154 patients enrolled in the cohort, however not all patients have complete PROMs data. Also due to illogically ordered dates 19 patients were dropped from our sample. We have baseline EQ-5D-3L scores for 1,076 patients, and EORTC-8D scores for 1,074 patients, the complete case sample (where there is a value for both instruments) is 1,062 patients. We have information at first follow-up (either three or six months) for over 550 patients, and we are able to estimate QALYs for 507 patients (complete case). Note that 130 patients (over 10% of the sample) have died.

2.2. Instruments

The EORTC QLQ-C30 is a non-preference based health related quality of life (HRQoL) measure which is frequently employed in cancer clinical trials. It is one of a suite of EORTC instruments, and is regarded as a 'generic' cancer measure (e.g. EORTC QLQ-BR23 is specific to breast cancer, while EORTC QLQ-MY20 is for myeloma).[7] It includes 30 questions, which feed into nine multi-item scales: five functional scales (physical, role, cognitive, emotional and social functioning); three symptom scales (fatigue, pain and nausea/Vomiting); and a global health status/quality of life scale. Six single-item scales mainly for symptoms are also included (dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties). There are three versions of the questionnaire, Cancer 2015 includes version 3.0. It provides summary scores (between 0 and 100) for a patient's functioning, symptoms and global quality of life; where a higher score represents a higher ("better") level of functioning or quality of life, or a higher ("worse") level of symptoms.

The EORTC-8D has eight dimensions (physical functioning, role functioning, pain, emotional functioning, social functioning, nausea, fatigue and sleep disturbance, and constipation/diarrhoea) and is derived using Rasch and factor analysis from the EORTC QLQ-C30.[8] Note that the original dataset of EORTC QLQ-C30 responses were from patients with multiple myeloma. There are 81,920 unique health states which were valued using a time trade-off approach in healthy individuals from the North of England. The resulting utility scores range from 0.292 to 1.00.

The EQ-5D-3L (previous known as the EQ-5D) has five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) each with three levels, such that there are 243 health states.[9] Note that there is now a five level version, EQ-5D-5L, which may be included in the next phase of Cancer 2015.[10] The EQ-5D-3L was also valued using a time trade-off approach in the United Kingdom (UK). Other country valuations exists, some of which use other valuation techniques (including a Discrete Choice Experiment (DCE) in Australia [11]), however, this analysis ignores any cross country differences, and given there only exists a UK tariff for the EORTC-8D, we

use the UK tariff (MVH-A1 algorithm) for the EQ-5D-3L.[12] The scoring range for the EQ-5D-3L is from -0.594 to 1.000, i.e. it includes states that are worse than dead (<0).

Cancer 2015 also included the EORTC-8D as an instrument in its own right, rather than as a mapping algorithm of EORTC QLQ-C30 responses. Although it has not been validated as an instrument (e.g. psychometrically assessed) it was thought that it would be useful to test the instrument relative to the algorithm as the instrument has ten questions (across eight domains) while the EORTC QLQ-C30 has 30 questions, thus we could test for 'efficiency' in questionnaire length. This current analysis ignores the responses of the 8D instrument, and instead uses the 8D algorithm responses. This is because they were highly correlated (r=0.93) and the 8D instrument, despite being shorter, had more missing responses (which was possibly due to an ordering effect in the questionnaire). Note that the next phase of Cancer 2015 will not include the EORTC-8D as an instrument.

2.3. Analysis

Initially we assessed the similarities of the instrument by comparing the domains, and the contribution that each domain makes to the overall utility score. This contribution was calculated by dividing the maximal decrease of a domain (the coefficient attached to the lowest/worst level) by the total decrease in utility for the instrument. The interaction terms, particularly the N3 term in the EQ-5D-3L algorithm were ignored in this calculation.

Next using the baseline utility scores we assessed the normalities of the distribution of each instrument, using both the Shapiro-Wilk W test and the Shapiro-Francia W test. Skewness and kurtosis were also assessed.

The ceiling effects of each instrument were assessed by considered the responses within each level of each domain, when patients reported no problems.

We then assessed the correlation both within domains and the utility scores as a whole. This provides an assessment of construct validity. We further explored this by considering the relationship between item domains in one instrument and perfect health in the other instrument as measured by a utility score of 1, e.g. EORTC-8D item responses when the EQ-5D-3L is one, and vice versa.

Agreement between the instruments was examined using a Bland-Altman plot.[13] We first standardised the instruments given they have different ranges, such that the utility scores for both the EQ-5D-3L and EORTC-8D were measured continuously on the 0 to 1 scale.[14] Agreement was further assessed by estimating the intra-class correlation coefficient (ICC).[15]

To understand if either instrument is sensitive (or indeed more sensitive) to different covariates we compared means in the utility scores using paired t tests and ANOVAs where appropriate.

QALYs were estimated using the area under the curve method. Correlation between the generic QALYs and condition-specific QALYs was assessed using Spearman's rank correlation coefficients and a visual assessment of the scatterplot. The sensitivity of QALYs to various covariates was also explored in bivariate analyses, and in a more complete multivariate regression model. The data were reshaped (wide to long) and a Chow test was used to test for subgroup differences according to the instrument employed to estimate utilities.[16] The regression includes baseline utility scores as an explanatory variable, as well as patient demographics, disease characteristics, and indicators of severity. Those censored by death were excluded in the QALY regression analysis.

All statistical analyses were undertaken in STATA MP version 13.0.

3. Results

Figure 1 presents at initial qualitative assessment of the instruments, irrespective of the sample responses. The coloured pie 'slices' show that there are some similarities between the instruments. All of the EQ-5D-3L domains feature in the EORTC-8D except self care. EORTC-8D's role functioning and social functioning have been deemed to be equivalent to the EQ-5D-3L's usual activities, as role refers to pursuing leisure activities and social refers to interference with social activities, i.e. activities that one would usually undertake. The proportions clearly differ (and must be by virtue of the fact that there are five domains relative to eight); the pain domain makes the greatest contribution to the EQ-5D-3L utility score, while the emotional domain (closely followed by the social domain) makes the greatest contribution to the EORTC-8D utility score.

Table 1 presents the descriptive statistics for the sample. The sample is relatively old (mean age 62), the majority are male (55%) and a large number have other co-morbidities as measured by the Charlson Index (mostly diabetes and arthritis).[17] The cohort purposely included a private hospital in the sample (in order to make treatment comparisons at a later date, and also because Australia has a two-tiered system); this hospital contributed 23% of the patients, but 46% of the total sample also have insurance cover for hospitals. This variable can be considered to be reflective of income, as at a certain income threshold private health insurance is incentivised (i.e. an additional tax is imposed on high income earners who don't have insurance).

In terms of disease, prostate cancer and breast cancer contribute the most patients (16% each) to the cohort, but there is generally good representation across the spectrum of tumour sites. When staged (via the staging method appropriate to the tumour site) most are considered to be

curative (82%), although some patients have palliative treatment intentions. It is important to note that an inclusion criterion was that patients (for the presenting malignancy) were treatment naïve. A large majority of patients (64%) have an Eastern Oncology Cooperative Group (ECOG) status which aligns with normal activity.[18]

The EORTC QLQ-C30 summary measures suggest that these patients have good functioning and quality of life, and few symptoms. The mean EQ-5D-3L utility score was 0.752, while the mean utility score for the EORTC-8D was 0.833. The condition-specific score is likely to be higher, simply because it has a higher minimum score (0.292) compared to the generic measure (-0.594). The range of utility scores can be seen in Figure 2 which plots the histograms for the baseline scores for each instrument. The data is clearly skewed and non-normal, this is further supported in formal statistical tests (EQ-5D-3L Shapiro–Wilk test $z = 10.8$, $p<0.001$; EORTC-8D Shapiro–Wilk test $z = 9.3$, $p<0.001$).

In terms of responses to each domain, considerable variability in both instruments is observed, the use of the highest level (no problems) in the EQ-5D-3L ranged from 91.9% of responses for usual activities to 50.1% of responses for pain. For the EORTC-8D the use of the highest level ranges from 77.8% of response for nausea and vomiting and 32.5% of responses for fatigue.

The construct validity of the instruments is assessed by considering the correlations across the domains and the utility scores. Table 2 shows that correlations between the domains of the EQ-5D-3L and the EORTC-8D are quite high, and all are significant at the $p<0.001$ level. Tables 3a and 3b similarly show good construct validity; when one instrument records a score of full perfect health for the most part this corresponds with the higher levels in each domain in the other instrument. The correlation between the baseline utility scores is 0.765, $0.7 < r < 0.9$ is considered to be strongly correlated. Figure 3 present this correlation graphically in a scatterplot. The correlations between the utility scores and the EORTC QLQ-C30 summary scores are also high, ranging from 0.738 to 0.919, except for the correlation between EQ-5D-5L utility score and the global quality of life summary score which is 0.673.

The ICC was 0.641 which suggests the agreement between the measures is good.[15] The Bland-Altman Plot in Figure 4 suggests that there are small mean differences, but relatively wide limits of agreement. Similar wide confidence intervals have been reported by others. [19, 20]

An analysis of the sensitivity of each instrument to various subgroups including patient and disease characteristics (see Table 4, columns 1-4) finds that both the EQ-5D-3L and the EORTC-8D are sensitive to gender (females have lower baseline utility scores), admitting hospital (public patients have lower utility scores), smoking status (smokers, including ex-smokers have lower utility

scores), stage of disease (metastatic cancer patients have lower utility scores), hospital insurance (those without insurance have lower utility scores), expected future follow-up (those with plans for follow-up at three months have lower baseline utility scores) and ECOG score (those with worse scores have lower utility scores). There is also variation in cancer site, but the instruments find that prostate cancer patients have the highest baseline utility scores, while patients with lung cancer and cancer of the unknown primary have the lowest baseline utility scores. These findings imply that both measures at baseline have good discriminatory power.

While it is important to understand how these two instruments compare at baseline, for economic evaluations, and specifically cost utility analyses, of more importance is how these instruments compare when used to estimate QALYs. The above analysis suggests that there are apparent differences (at least in the scale) in the EQ-5D-3L and EORTC-8D at baseline, it is necessary to understand if these differences in scale translate to differences in QALYs.

The estimated mean QALYs when using the EQ-5D-3L is 0.560 (range of 0.040-1.444), the estimated mean QALYs when using the EORTC-8D is 0.589 (range 0.118-1.261), thus the QALY estimates are higher for the condition-specific measure and the range is narrower, the difference while small (0.029) is statistically significant at the margin ($p=0.076$, paired t test). The generic and condition-specific QALYs are highly correlated (Spearman's $r=0.915$), also see Figure 5.

The sensitivity of both types of QALYs to variations in the sample is presented in Table 4, columns 5-8. Similar relationships are found as reported for baseline utility scores, although there is now no difference in terms of smoking status for either instrument's QALYs, while the generic QALYs (those estimated using the EQ-5D-3L) are insensitive to the stage of disease. To directly compare the QALYs a multivariate seemingly unrelated regression estimation using robust standard errors was undertaken; the results of this analysis are reported in Table 5. Few significant covariates are apparent (baseline utility and being treated in a private hospital have positive effects on QALYs; colorectal cancer relative to other cancers, worse ECOG status, and a decline in ECOG status have negative effects on QALYs), and Chow tests comparing the instruments finds that there are no statistically significant differences in the coefficients within each equation. The lack of statistically significant difference implies that the difference between the generic and condition-specific QALYs is not driven by any of these covariates.

4. Discussion

The health economics discipline has been debating condition-specific measures in the literature for a number of years.[2, 21-23] Recently there has been a plethora of condition-specific measures being developed,[24-26] but their use in decision making remains limited. This paper

seeks to further inform the debate by testing the validity, responsiveness and sensitivity of a CSPBM for cancer. The EORTC-8D has previously been found to be broadly comparable to the EQ-5D,[27] but that was within the same dataset that the EORTC-8D was developed from, hence this study provides the first external assessment of the instrument.

At baseline the two instruments were found to be well correlated and have good agreement, they were equally as sensitive to a range of patient and disease characteristics. The utility score for the EORTC-8D was higher than for the EQ-5D-3L, but this is likely to be a function of the EORTC-8D have a higher 'floor', the lowest possible utility score is 0.292 compared to the EQ-5D-3L floor of -0.594.

The QALY estimates were also highly correlated, but notably the condition-specific QALYs estimated using the EORTC-8D were significantly higher than those derived from differences in the EQ-5D-3L over time. Both the generic and condition-specific QALYs were found to be sensitive to a number of patient and disease characteristics, but equally sensitive, such that we failed to identify why the condition-specific QALYs were higher. The obvious answer is that the EORTC-8D is superior to the generic EQ-5D-3L when it comes to measuring health status in cancer patients.

Economic evaluations of cancer care, particularly in relation to personalised medicine are central research issues.[28, 29] Much of the concern is with regard to the high cost of these drugs (and test and treatment combinations), but it is equally important to correctly measure the outcomes of treatment (and testing). Our analysis and previous research[27] might imply that CSPBM do this better than generic measures, but we may be premature in making this judgement, we also don't know if 'better' necessarily equates to higher QALYs (which would be required if we wish to trade-off some of the additional costs of cancer drugs). The QALY difference identified in this analysis, 0.029, means that if the end-of-life/cancer cost effectiveness threshold were around £50,000 then the cost difference using the two instruments could be as much as £1300; simplistically treatment could cost £1300 more and decision makers would be willing to accept this if condition-specific QALYs are used.

One limitation of this study is that while the cohort is rich in information, it is not a clinical trial and therefore treatment effects vary. More research is required to compare the generic PBM and CSPBM in a trial setting. A further future limitation is that an additional PBM using the EORTC QLQ-C30 is under-development (MAUCa Consortium). This utilises data from patients receiving palliative radiotherapy for recurrent or metastatic cancer from various primary sites. The existence of another CSPBM for cancer using the EORTC QLQ-C30 will further muddy the waters, and potentially lead to gaming by pharmaceutical manufacturers who are seeking public reimbursement. The benefit may in fact become a burden.[22]

A number of methodological issues with regard to the use of QALYs in cancer remain outstanding. [30] The valuation of the EORTC-8D employed a time trade-off approach (as did the EQ-5D-3L), which arguably requires assumptions that are violated in end of life scenarios. Also (again as with the EQ-5D-3L) members of the public were used in the valuation study, and it is argued that the general public may have misunderstandings of what it is to live with cancer. Despite these limitations, the EORTC-8D does appear to be more sensitive than the EQ-5D-3L, and this research provides further evidence for decision makers who may wish to consider CSPBM.

Points for HESG Discussion:

- Is it useful to focus on the cohort as a whole, or would it be informative to consider cross-sections of specific tumour streams?
- What further analyses could be undertaken, particularly beyond the baseline utility values, so focusing on more QALYs? Is it worth looking at parts of the distribution, the more variable QALYs?
- What should we do with those who have died (currently excluded in QALY estimation)? Take their utility to 0 at time of death and estimate QALYs?
- How can we validate the EORTC-8D against a clinical measure? The EORTC QLQ-C30 is used to produce the EORTC-8D. Can we use changing treatment regime (once pharma drug are linked)?

References

1. National Institute for Care and Excellence. Guide to the Methods of Technology Appraisal. 2013 30 November 2013]; Available from: http://publications.nice.org.uk/guide-to-the-methods-of-technology-appraisal-2013-pmg9/foreword.
2. Brazier JDixon S, The Use of Condition Specific Outcome Measures in Economic Appraisal. *Health economics*, 1995. **4**(4): p. 255-64.
3. Petrillo JCairns J, Converting Condition-Specific Measures into Preference-Based Outcomes for Use in Economic Evaluation. *Expert review of pharmacoeconomics & outcomes research*, 2008. **8**(5): p. 453-61.
4. Brazier JE, Yang Y, Tsuchiya A, Rowen DL, A Review of Studies Mapping (or Cross Walking) Non-Preference Based Measures of Health to Generic Preference-Based Measures. *The European journal of health economics : HEPAC : health economics in prevention and care*, 2010. **11**(2): p. 215-25.
5. Stevens K, Developing a Descriptive System for a New Preference-Based Measure of Health-Related Quality of Life for Children. *Quality of Life Research*, 2009. **18**(8): p. 1105-13.
6. Brazier J, Roberts J, Deverill M, The Estimation of a Preference-Based Measure of Health from the Sf-36. *Journal of health economics*, 2002. **21**(2): p. 271-92.
7. Cella DF, Wiklund I, Shumaker SA, Aaronson NK, Integrating Health-Related Quality of Life into Cross-National Clinical Trials. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 1993. **2**(6): p. 433-40.
8. Rowen D, Brazier J, Young T, Gaugris S, Craig BM, King MT, Velikova G, Deriving a Preference-Based Measure for Cancer Using the Eortc Qlq-C30. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 2011. **14**(5): p. 721-31.
9. Brooks R, Euroqol: The Current State of Play. *Health policy*, 1996. **37**(1): p. 53-72.
10. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X, Development and Preliminary Testing of the New Five-Level Version of Eq-5d (Eq-5d-5l). *Quality of Life Research*, 2011. **20**(10): p. 1727-36.
11. Viney R, Norman R, Brazier J, Cronin P, King MT, Ratcliffe J, Street D, An Australian Discrete Choice Experiment to Value Eq-5d Health States. *Health economics*, 2013.
12. Dolan P, Modeling Valuations for Euroqol Health States. *Medical care*, 1997. **35**(11): p. 1095-108.
13. Bland MJAltman DG, Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *The lancet*, 1986. **327**(8476): p. 307-10.
14. Seymour J, McNamee P, Scott A, Tinelli M, Shedding New Light onto the Ceiling and Floor? A Quantile Regression Approach to Compare Eq-5d and Sf-6d Responses. *Health economics*, 2010. **19**(6): p. 683-96.
15. Rosner BA, The Intraclass Correlation Coefficient, in *Fundamentals of Biostatistics*. 2011, Duxbury: Pacific Grove, CA.
16. Chow GC, Tests of Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica: Journal of the Econometric Society*, 1960: p. 591-605.
17. Charlson ME, Pompei P, Ales KL, MacKenzie CR, A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of chronic diseases*, 1987. **40**(5): p. 373-83.

18. Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, Carbone PP, Toxicity and Response Criteria of the Eastern Cooperative Oncology Group. *American journal of clinical oncology*, 1982. **5**(6): p. 649-56.

19. Turner N, Campbell J, Peters TJ, Wiles N, Hollinghurst S, A Comparison of Four Different Approaches to Measuring Health Utility in Depressed Patients. *Health and quality of life outcomes*, 2013. **11**.

20. van Stel HFBuskens E, Comparison of the Sf-6d and the Eq-5d in Patients with Coronary Heart Disease. *Health and quality of life outcomes*, 2006. **4**(1): p. 20.

21. Dowie J, Decision Validity Should Determine Whether a Generic or Condition-Specific Hrqol Measure Is Used in Health Care Decisions. *Health economics*, 2002. **11**(1): p. 1-8.

22. Versteegh MM, Leunis A, Uyl-de Groot CA, Stolk EA, Condition-Specific Preference-Based Measures: Benefit or Burden? *Value in Health*, 2012. **15**(3): p. 504-13.

23. Brazier JFitzpatrick R, Measures of Health-Related Quality of Life in an Imperfect World: A Comment on Dowie. *Health economics*, 2002. **11**(1): p. 17-9; discussion 21-2.

24. Yang Y, Brazier J, Tsuchiya A, Coyne K, Estimating a Preference-Based Single Index from the Overactive Bladder Questionnaire. *Value in Health*, 2009. **12**(1): p. 159-66.

25. Brazier JE, Roberts J, Platts M, Zoellner YF, Estimating a Preference-Based Index for a Menopause Specific Health Quality of Life Questionnaire. *Health and quality of life outcomes*, 2005. **3**(13): p. 3-13.

26. Yang Y, Brazier JE, Tsuchiya A, Young TA, Estimating a Preference-Based Index for a 5-Dimensional Health State Classification for Asthma Derived from the Asthma Quality of Life Questionnaire. *Medical decision making : an international journal of the Society for Medical Decision Making*, 2011. **31**(2): p. 281-91.

27. Rowen D, Young T, Brazier J, Gaugris S, Comparison of Generic, Condition-Specific, and Mapped Health State Utility Values for Multiple Myeloma Cancer. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, 2012. **15**(8): p. 1059-68.

28. Annemans L, Redekop K, Payne K, Current Methodological Issues in the Economic Assessment of Personalized Medicine. *Value in Health*, 2013. **16**(6): p. S20-S6.

29. Jönsson B, Technology Assessment for New Oncology Drugs. *Clinical Cancer Research*, 2013. **19**(1): p. 6-11.

30. Garau M, Shah KK, Mason AR, Wang Q, Towse A, Drummond MF, Using Qalys in Cancer. *PharmacoEconomics*, 2011. **29**(8): p. 673-85.

Figure 1a: Maximal theoretical contributions of the EQ-5D-3L

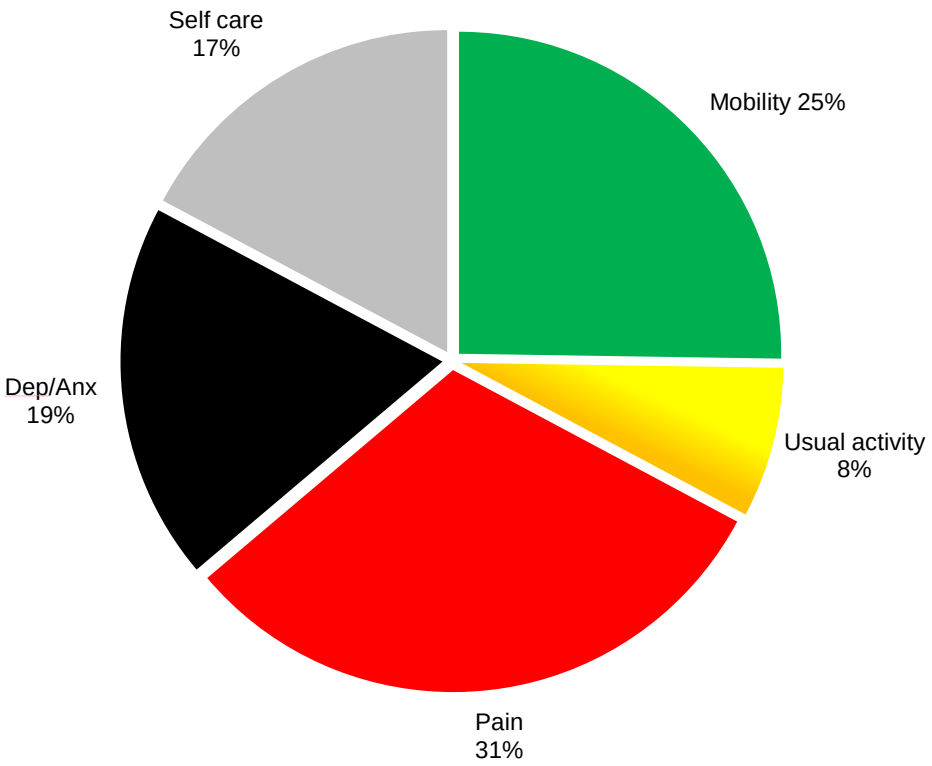

Figure 1b: Maximal theoretical contributions of the EORTC-8D

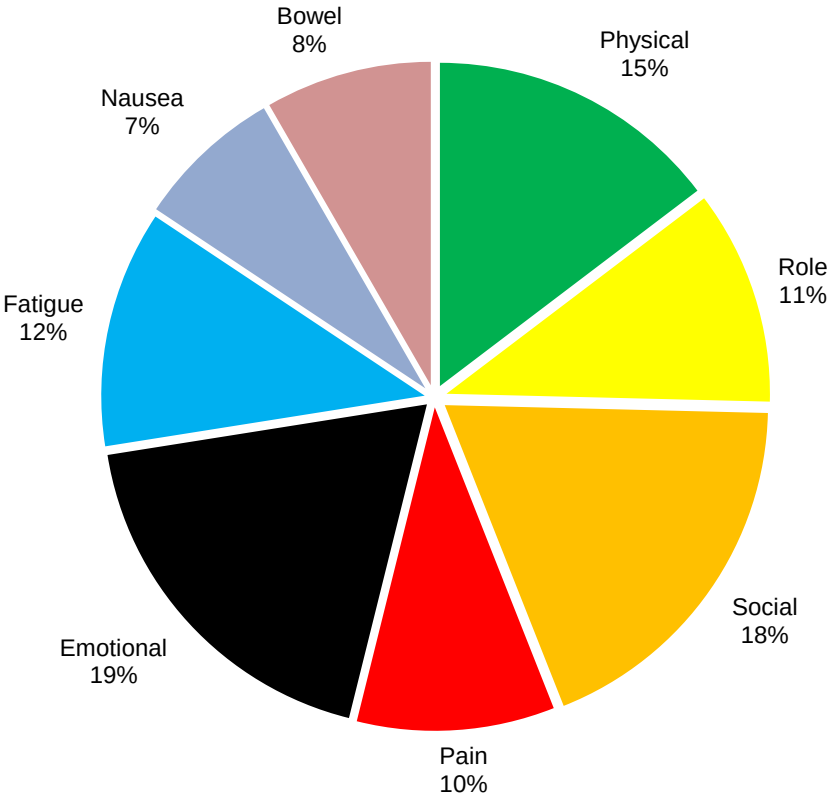

Figure 2: Histogram of baseline EQ-5D and EORTC-8D scores

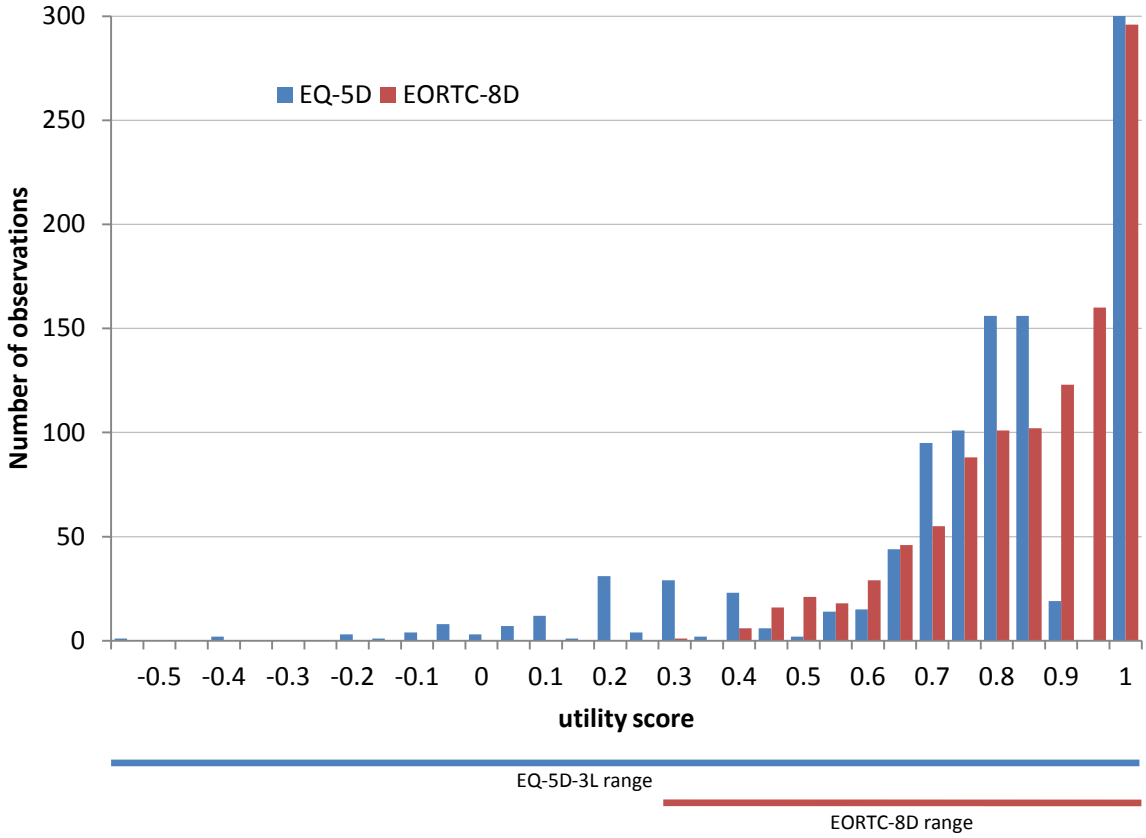

Figure 3: Correlation in baseline utility scores

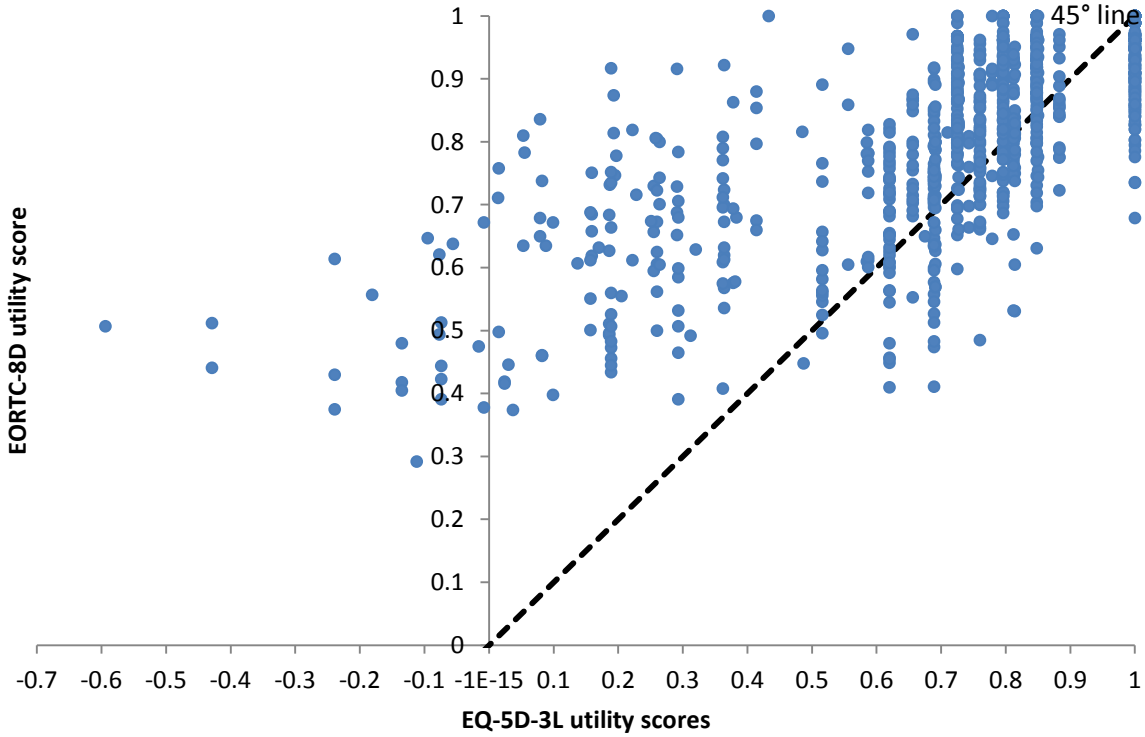

Figure 4: Bland Altman plot of EQ-5D-3L and EORTC-8D at baseline

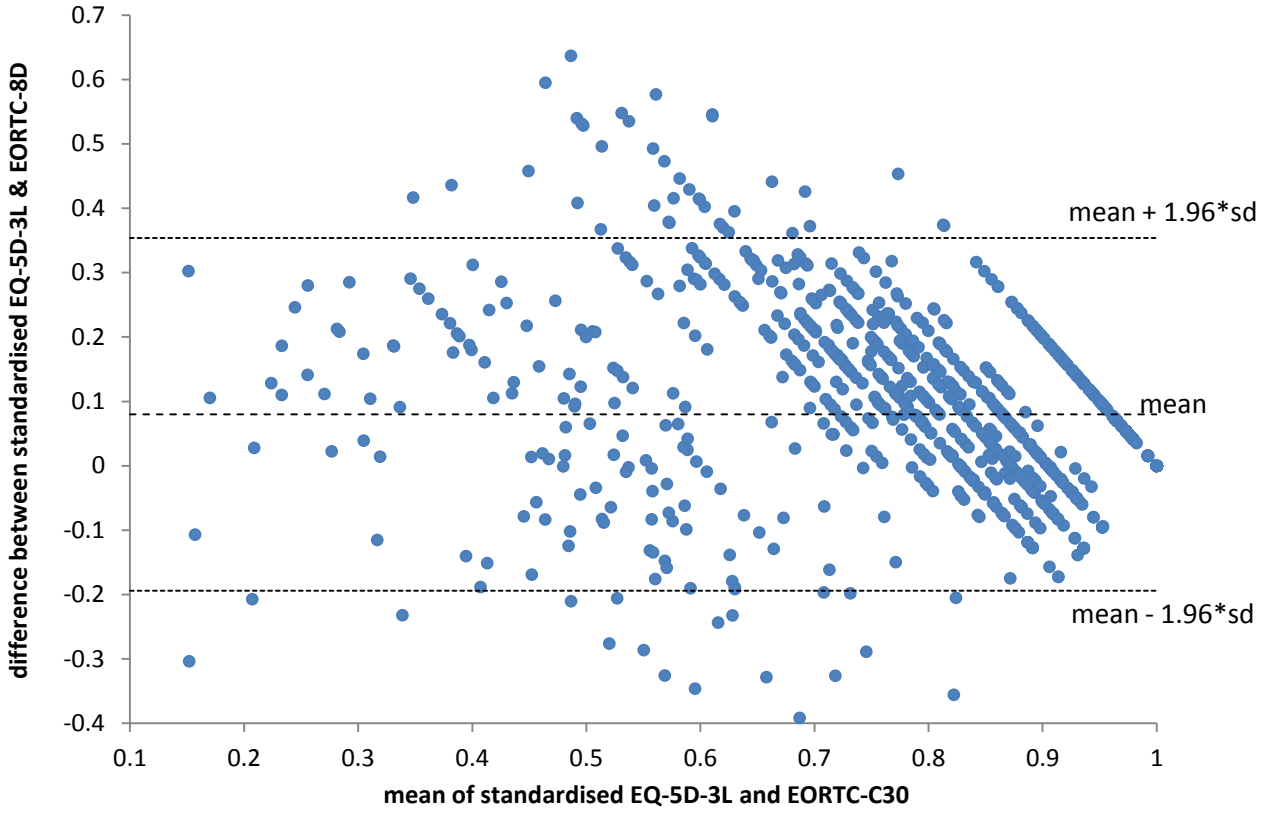

Figure 5: Correlation in QALY estimates

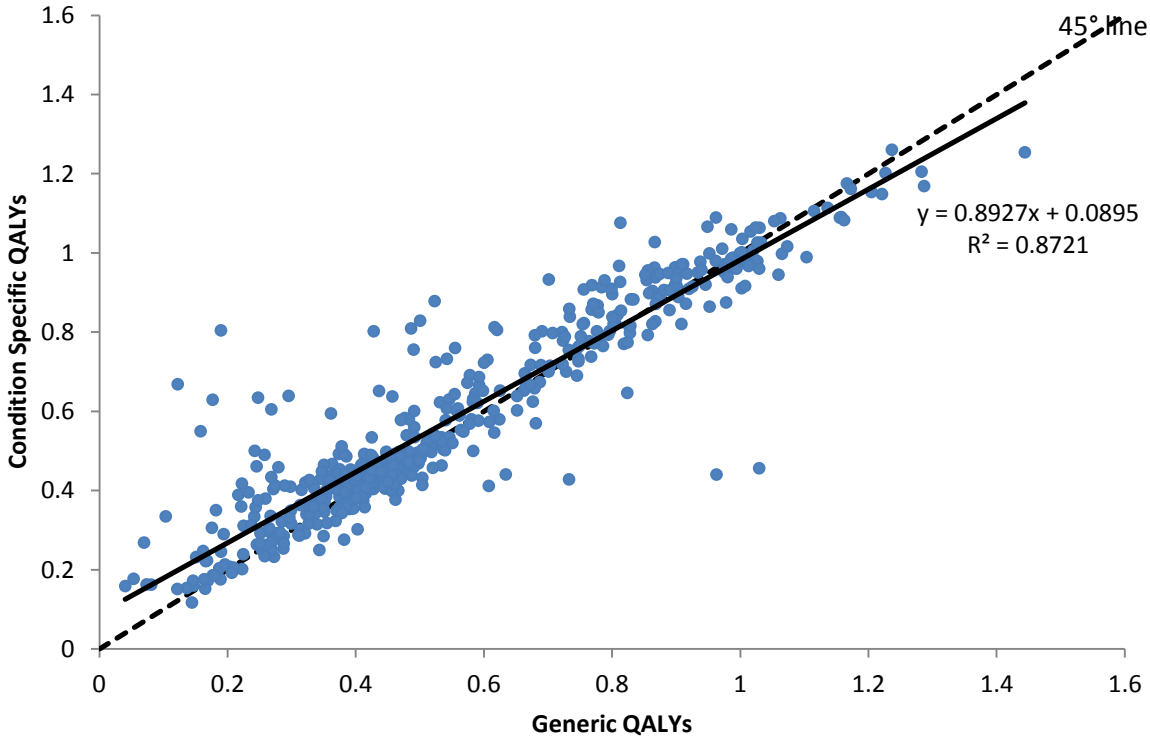

Table 1: Baseline sample descriptive statistics (N=1062)

		Mean (range) or Proportion
Age at consent		62 (18, 92)
Charlson Comorbidity Index (N=944)		2.22 (0, 14)
Gender	Female	44.9%
	Male	55.1%
Recruiting hospital	Public	76.3%
	Private	23.7%
Hospital insurance cover (N=1042)	Yes	45.68%
	No	54.32%
Smoking status (N=1031)	Current smoker	15.9%
	Ex-smoker	44.3%
	Never smoked	39.8%
Residence (N=1049)	Major city	48.4%
	Inner regional	45.6%
	Outer regional	6.0%
Tumour site (N=1059)	Prostate	15.8%
	Breast	15.9%
	Head and Neck	12.6%
	Colorectal	11.0%
	Lung	10.6%
	Bone and Soft Tissue	5.1%
	Cervical	3.9%
	Cancer Unknown Primary	3.4%
	Renal	3.4%
	Oesophagogastric	3.4%
	Other (includes 12 known sites)	15.1%
Treatment intentions (N=976)	No treatment	1.84%
	Curative	81.86%
	Palliative	16.29%
ECOG score (N=1040)	Normal activity	63.8%
	Limited in normal activity	24.8%
	Self care capable but not working	8.5%
	Limited self care	2.7%
	No self-care	0.3%
EQ-5D-3L		0.752 (-0.59, 1.00)
EORTC-8D		0.833 (0.29, 1.00)
C30 Functioning Score		79.48 (8.88, 100)
C30 Symptom Score		18.88 (0, 87.18)
C30 Global Health Score		69.52 (0, 100)

Table 2: Correlations between utility domains at baseline

EORTC-8D	Mobility	Self Care	EQ-5D-3L Usual Activities	Pain	Anxiety/ Depression
Physical Functioning	**0.608**	0.362	0.530	0.435	0.234
Role Functioning	0.436	0.357	**0.651**	0.444	0.265
Pain	0.436	0.297	0.524	**0.613**	0.267
Emotional Functioning	0.195	0.152	0.282	0.289	**0.645**
Social Functioning	0.378	0.334	**0.601**	0.425	0.347
Fatigue and sleep	0.346	0.282	0.494	0.432	0.365
Nausea	0.273	0.251	0.420	0.364	0.280
Constipation/ Diarrhoea	0.290	0.224	0.338	0.297	0.230

Note: all correlations are significant at the $p<0.01$ level; correlations between like domains are in bold.

Table 3a: EORTC-8D responses when EQ-5D-3L=1

	Level 1	Level 2	Level 3	Level 4	Level 5
Physical Functioning	80.80	16.41	2.48	0.31	0.00
Role Functioning	90.40	8.36	0.31	0.93	n/a
Pain	93.50	5.88	0.31	0.31	n/a
Emotional Functioning	88.24	11.46	0.31	0.00	n/a
Social Functioning	86.38	12.38	1.24	0.00	n/a
Fatigue and sleep	62.54	33.13	4.02	0.31	n/a
Nausea	96.59	3.10	0.31	0.00	n/a
Constipation/Diarrhoea	82.04	14.86	2.17	0.93	n/a

Table 3b: EQ-5D-5L responses when EORTC-8D=1

	Level 1	Level 2	Level 3
Mobility	98.63	1.37	0.00
Self-care	100.00	0.00	0.00
Usual activities	99.32	0.68	0.00
Pain	92.47	7.53	0.00
Anxiety/Depression	93.15	6.85	0.00

Table 4: Differences in baseline utility and QALYs

	EQ-5D-3L baseline utility		EORTC-8D baseline utility		Generic QALYs		Condition-specific QALYs	
	mean	p-value	mean	p-value	mean	p-value	mean	p-value
Male	0.775	0.002	0.852	<0.001	0.594	<0.001	0.626	<0.001
Female	0.724		0.810		0.509		0.535	
Public hospital	0.733	<0.001	0.822	<0.001	0.526	<0.001	0.562	<0.001
Private hospital	0.814		0.87		0.664		0.675	
Planned six month follow-up	0.793	0.001	0.859	<0.001	-	-	-	-
Planned three month follow-up	0.717		0.789		-		-	
Hospital insurance – no	0.716	<0.001	0.815	<0.001	0.528	0.009	0.561	0.013
Hospital insurance – yes	0.796		0.856		0.591		0.618	
Smoker	0.704	0.014	0.818	0.028	0.547	0.115	0.59	0.193
Ex-smoker	0.748		0.825		0.535		0.566	
Never smoked	0.775		0.848		0.587		0.611	
ECOG – Normal activity	0.844	<0.001	0.893	<0.001	0.601	<0.001	0.622	<0.001
ECOG – Limited in normal activity	0.676		0.779		0.503		0.54	
ECOG – Self care capable but not working	0.504		0.660		0.384		0.455	
ECOG - Limited self care	0.212		0.545		0.368		0.454	
ECOG - No self-care	0.190		0.689		-			
Change in ECOG – none	-	-	-	-	0.628	<0.001	0.653	<0.001
Change in ECOG – decline	-		-		0.478		0.511	
Change in ECOG – improvement	-		-		0.534		0.579	
Stage - unknown	0.786	<0.001	0.846	<0.001	0.513	0.102	0.551	0.005
Stage - metastases	0.572		0.741		0.486		0.499	
Stage - regional lymph nodes	0.756		0.827		0.537		0.574	
Stage - invasion of adjacent	0.728		0.812		0.505		0.54	
Stage - localised	0.798		0.862		0.601		0.629	
Aged < 30	0.535	0.043	0.582	0.022	0.535	0.033	0.582	0.009
Aged 30-50	0.542		0.566		0.526		0.566	
Aged 50-70	0.603		0.616		0.589		0.619	
Aged >70	0.533		0.524		0.509		0.525	
Treatment intent – none	0.740	<0.001	0.822	<0.001	0.463	0.005	0.488	0.001
Treatment intent – curative	0.791		0.857		0.58		0.609	
Treatment intent – palliative	0.556		0.718		0.45		0.47	
Alive at follow-up	0.779	<0.001	0.847	<0.001	-		-	
Dead at follow-up	0.551		0.728		-		-	
Site – prostate	0.888	<0.001	0.930	0.001	0.673	<0.001	0.697	<0.001

Site – breast	0.773	0.841	0.541	0.561
Site – head & neck	0.736	0.834	0.546	0.593
Site – colorectal	0.810	0.823	0.515	0.495
Site – lung	0.726	0.801	0.519	0.551
Site – bone & soft tissue	0.753	0.855	0.539	0.618
Site – cervical	0.800	0.863	0.549	0.585
Site – CUP	0.693	0.811	0.384	0.437
Site – renal	0.745	0.842	0.473	0.529
Site – oesophagogastric	0.745	0.835	0.429	0.473
Site – all other	0.735	0.814	0.557	0.586

Table 5: QALY regression results

	coefficient	robust std err	p-value
EQ-5D QALYs			
Baseline utility	0.341	0.048	0.000
Age	0.000	0.001	0.972
Female	-0.043	0.031	0.165
Private Hospital	0.165	0.036	0.000
Smoker	-0.017	0.036	0.625
Ex-smoker	-0.027	0.025	0.295
No hospital insurance	-0.030	0.026	0.258
Prostate	-0.019	0.049	0.698
Breast	-0.040	0.048	0.403
Head and Neck	0.043	0.047	0.355
Colorectal	-0.100	0.046	0.030
Lung	0.064	0.064	0.314
Bone and Soft Tissue	0.047	0.058	0.421
Cervical	0.008	0.061	0.902
Cancer of the Unknown Primary	-0.130	0.105	0.214
Renal	-0.031	0.093	0.742
Oesophagogastric	-0.106	0.072	0.139
Limited in normal activity	-0.114	0.035	0.001
Self-care capable but not working	-0.154	0.055	0.005
Limited self-care, 50% disabled	-0.062	0.123	0.611
ECOG status did not change	-0.053	0.043	0.219
ECOG status declined	-0.199	0.045	0.000
Curative treatment plan	0.061	0.113	0.593
Palliative treatment plan	0.038	0.119	0.749
Unknown stage	-0.118	0.067	0.076
Metastases	0.028	0.052	0.594
Regional lymph nodes	-0.003	0.031	0.915
Invasion of adjacent	-0.007	0.037	0.843
Constant	0.433	0.161	0.007
log(variance)	-2.888	0.054	0.000

EORTC-8D QALYs			
Baseline utility	0.463	0.093	0.000
Age	-0.001	0.001	0.540
Female	-0.031	0.031	0.315
Private Hospital	0.133	0.034	0.000
Smoker	-0.010	0.035	0.773
Ex-smoker	-0.021	0.025	0.385
No hospital insurance	-0.022	0.025	0.366
Prostate	0.010	0.049	0.832
Breast	-0.029	0.049	0.558
Head and Neck	0.052	0.048	0.274
Colorectal	-0.113	0.046	0.015
Lung	0.065	0.062	0.298
Bone and Soft Tissue	0.068	0.055	0.223
Cervical	0.000	0.060	0.998
Cancer of the Unknown Primary	-0.123	0.098	0.211
Renal	-0.037	0.087	0.669
Oesophagogastric	-0.081	0.079	0.302
Limited in normal activity	-0.087	0.035	0.014
Self-care capable but not working	-0.091	0.054	0.090
Limited self-care, 50% disabled	-0.002	0.119	0.989
ECOG status did not change	-0.043	0.043	0.318
ECOG status declined	-0.179	0.045	0.000
Curative treatment plan	0.098	0.114	0.391
Palliative treatment plan	0.073	0.121	0.547
Unknown stage	-0.102	0.057	0.075
Metastases	0.022	0.052	0.680
Regional lymph nodes	0.009	0.030	0.779
Invasion of adjacent	-0.013	0.036	0.713
Constant	0.281	0.184	0.128
log(variance)	-2.931	0.049	0.000